

## Introduction

IMG ER 3.5 contains RNASeq data from *Synechococcus* study. This data can be explored using the tools illustrated in Figures 1 and 2.

**Synechococcus sp. PCC 7002**

[Browse Genome](#) [BLAST Genome](#) [Download Data](#)

**About Genome**

- [Overview](#)
- [Statistics](#)
- [Expression Studies](#)
- [Genes](#)

**Overview**

<b>Proposal Name</b>	
<b>Organism Name</b>	Synechococcus sp. PCC 7002
<b>Taxon Object ID</b>	641522654
<b>NCBI Taxon ID</b>	<a href="#">32049</a>
<b>GOLD ID in IMG Database</b>	<a href="#">Project Id: Gc00746</a>
<b>External Links</b>	<a href="#">JGI Portal</a> ; <a href="#">NCBI/RefSeq:NC_010474</a> ; <a href="#">NCBI/RefSeq:NC_010477</a> ; <a href="#">NCBI/RefSeq:NC_010478</a> ; <a href="#">NCBI/RefSeq:NC_010479</a>
<b>Genome Type</b>	isolate

**Expression Studies (i)**

[List of RNASeq Studies](#)

**RNASeq Expression Studies (ii)**

[Synechococcus sp. PCC 7002](#)

Study ID	Study Name	Expressed Genes
3	<a href="#">Transcription profiling of Synechococcus sp. Strain PCC 7002</a>	2823

**RNASeq Study (iii)**

[Synechococcus sp. PCC 7002](#)

[Select Samples](#) [View In](#) [Multiple Sample Analysis](#)

**Study Description**

**Transcription profiling of Synechococcus sp. Strain PCC 7002**

**Synopsis:** Transcription profiling of the model cyanobacterium *Synechococcus* sp. strain PCC 7002 by Next-Gen (SOLID) sequencing of cDNA under different conditions and time points, with a total of 11 samples.

**Publication:** [PubMed](#)

**Contacts:**

**hint:** You may change any given sample

Select	Sample ID	Sample Name	Genes with Reads	Total Reads Count	Average Reads per Gene
<input type="checkbox"/>	SRR097642	<a href="#">dark aerobic</a>	2823	481909	170.71
<input type="checkbox"/>	SRR097643	<a href="#">dark anaerobic</a>	2823	1033135	365.97
<input type="checkbox"/>	SRR097644	<a href="#">high light</a>	2823	473131	167.6
<input type="checkbox"/>	SRR097645	<a href="#">standard1</a>	2823	3005584	1064.68
<input type="checkbox"/>	SRR097646	<a href="#">standard2</a>	2823	1470994	521.07
<input type="checkbox"/>	SRR097647	<a href="#">standard3</a>	2823	1485242	526.12
<input type="checkbox"/>	SRR097648	<a href="#">QD 0.4</a>	2823	498970	176.75
<input type="checkbox"/>	SRR097649	<a href="#">QD 1.0</a>	2823	859775	304.56
<input type="checkbox"/>	SRR097650	<a href="#">QD 3.0</a>	2823	520606	184.42
<input type="checkbox"/>	SRR097651	<a href="#">QD 5.0</a>	2823	375423	132.99
<input type="checkbox"/>	SRR097656	<a href="#">low Q2</a>	2823	407962	144.51

[Select All](#) [Clear All](#)

**Figure 1.** Organism Details – RNASeq Expression Studies

The **Organism Details** for a genome associated with rnaseq expression studies provides two starting points for exploring rnaseq expression data. First, an **Expression Studies** link, illustrated in Figure 1(i), leads to the list of studies for the genome of interest, as illustrated in Figure 1(ii). Each rnaseq expression study is displayed together with the number of expressed genes. The **Study Name** provides a link to the list of experiments/samples for the study, as illustrated in Figure 1(iii). The description for each sample consists of the experimental conditions and provides a link to the expression data for the sample organized per expressed gene, as illustrated in Figure 1(iv).

### Synechococcus sp. PCC 7002

[Browse Genome](#)
[BLAST Genome](#)
[Download Data](#)

#### About Genome

- [Overview](#)
- [Statistics](#)
- [Expression Studies](#)
- [Genes](#)

#### Overview (i)

Proposal Name	Synechococcus sp. PCC 7002
Organism Name	Synechococcus sp. PCC 7002
Taxon Object ID	641522654

---

#### Genome Statistics

Revised Genes	2	0.06%
Genes with RNASeq data	2823	
COG clusters	1416	29.06%

#### Gene Detail

- [Gene Information](#)
- [Find Candidate Product Name](#)
- [Evidence For Function Predictions](#)
- [Sequence Search](#)
- [External Sequence Search](#)
- [IMG Sequence Search](#)
- [Homolog Display](#)

#### Gene Information (ii)

Gene Information	
Gene Object ID	641610019
Gene Symbol	manA
Locus Tag	SYNPCC7002
Product Name	Mannose-6-ph
Protein Information	
Amino Acid Sequence Length	113aa
COG	COG0662 - M
COG Function	Carbohydrate
Families	<ul style="list-style-type: none"> <li>- HMMPanthe</li> <li>- HMMPanthe</li> <li>- IPR001538</li> <li>- IPR011051</li> <li>- IPR014710</li> </ul>
Transmembrane Helices	No
Signal Peptide	No
Statistics	<a href="#">peptide</a>
Expression	<a href="#">rnaseq</a>
Pfam	<a href="#">pfam01050</a> - MannoseP_isomer

#### RNASeq Data for Gene (iii)

>Gene: [641610019](#) mannose-6-phosphate isomerase, type 2 (EC 5.3.1.8) (IMGterm:M) [Synechococcus sp. PCC 7002: NC\_010475]

Filter column:  Filter term:

[Export](#) Page 1 of 1 << first < prev 1 next > last >> [All](#)

Sample ID	Description	Reads Count	Study
SRR097642	dark aerobic	14	<a href="#">Transcription profiling of Synechococcus sp. Strain PCC 7002</a>
SRR097643	dark anaerobic	41	<a href="#">Transcription profiling of Synechococcus sp. Strain PCC 7002</a>
SRR097644	high light	26	<a href="#">Transcription profiling of Synechococcus sp. Strain PCC 7002</a>
SRR097645	standard1	81	<a href="#">Transcription profiling of Synechococcus sp. Strain PCC 7002</a>
SRR097646	standard2	49	<a href="#">Transcription profiling of Synechococcus sp. Strain PCC 7002</a>
SRR097647	standard3	45	<a href="#">Transcription profiling of Synechococcus sp. Strain PCC 7002</a>
SRR097648	OD 0.4	24	<a href="#">Transcription profiling of Synechococcus sp. Strain PCC 7002</a>
SRR097649	OD 1.0	35	<a href="#">Transcription profiling of Synechococcus sp. Strain PCC 7002</a>
SRR097650	OD 3.0	20	<a href="#">Transcription profiling of Synechococcus sp. Strain PCC 7002</a>
SRR097651	OD 5.0	20	<a href="#">Transcription profiling of Synechococcus sp. Strain PCC 7002</a>
SRR097656	low O2	13	<a href="#">Transcription profiling of Synechococcus sp. Strain PCC 7002</a>

[Export](#) Page 1 of 1 << first < prev 1 next > last >> [All](#)

**Figure 2.** Gene Details - RNASeq Expression Data.

The **Organism Details** provides another starting point for exploring rnaseq expression data through the number of expressed genes in the Genome Statistics section, as illustrated in Figure 2(i). This number leads to the list of expressed genes, with each gene linked to its **Gene Detail** page, as illustrated in Figure 2(ii).

The **Protein Information** section of the **Gene Detail** for an expressed gene provides a link to **RNASeq Data**, as illustrated in Figure 2(iii), which displays the list of experiments/samples involving the expressed gene. Each sample is associated with a description consisting of the experimental conditions and the reads count for this experiment.

## Sample Analysis

The samples for each RNASeq study can be analyzed in a number of ways (Figure 3). Samples can be selected to perform single, pairwise, or multiple sample analysis. Samples can also be analyzed in GBrowse to view differences in read coverage for individual genes under various conditions.

### RNASeq Study ?

[Synechococcus sp. PCC 7002](#)

Select Samples
View in GBrowse
Single Sample Analysis
Pairwise Sample Analysis
Multiple Sample Analysis

**hint:** You may change the [Normalization Method](#) used to compute abundances of the genes in any given sample: ☐ Quantile ☐ Affine ☒ RPKM (Reads Per Kilobase of gene)

Column Selector

Select	Sample ID ▲	Sample Name	Genes with Reads	Total Reads Count	Average Reads per Gene
<input type="checkbox"/>	SRR097642	<a href="#">dark aerobic</a>	2823	481909	170.71
<input type="checkbox"/>	SRR097643	<a href="#">dark anaerobic</a>	2823	1033135	365.97
<input type="checkbox"/>	SRR097644	<a href="#">high light</a>	2823	473131	167.6
<input type="checkbox"/>	SRR097645	<a href="#">standard1</a>	2823	3005584	1064.68
<input type="checkbox"/>	SRR097646	<a href="#">standard2</a>	2823	1470994	521.07
<input type="checkbox"/>	SRR097647	<a href="#">standard3</a>	2823	1485242	526.12
<input type="checkbox"/>	SRR097648	<a href="#">OD 0.4</a>	2823	498970	176.75
<input type="checkbox"/>	SRR097649	<a href="#">OD 1.0</a>	2823	859775	304.56
<input type="checkbox"/>	SRR097650	<a href="#">OD 3.0</a>	2823	520606	184.42
<input type="checkbox"/>	SRR097651	<a href="#">OD 5.0</a>	2823	375423	132.99
<input type="checkbox"/>	SRR097656	<a href="#">low O2</a>	2823	407962	144.51

Select All
Clear All

**Figure 3.** RNASeq Study – Samples

## Single Sample Analysis

A single sample can be analyzed to show gene coverage and to display information and links to cassettes, associated cog functions, kegg pathways and modules, and EC (Figure 5). This information is displayed when the user selects a sample and clicks on “Gene Expression Summary” (Figure 4).

RNASeq Study ?

[Synchococcus sp. PCC 7002](#)

Select Samples

View in GBrowse

Single Sample Analysis

Pairwise Sample Analysis

Multiple Sample Analysis

Describe Sample

You may select a sample to compare and describe.

☐ Use only genes from gene cart

Gene Expression Summary

Pathways

**Figure 4.** Single Sample Analysis

The user can choose to view the chromosome for this organism colored by expression values for the selected sample by clicking on “Chromosome Viewer”.

RNASeq Expression Data for Selected Sample

[dark aerobic](#)

Chromosome Viewer

[Transcription profiling of Synchococcus sp. Strain PCC 7002](#)

Add Selections To Gene Cart

Select All

Clear All

Filter column: Gene ID Filter term:

Export

Page 1 of 29

<< first < prev

1

2

3

4

5

6

7

8

9

10

next > last >>

100

Column Selector

Select Page

Deselect Page

Select	Gene ID	Locus Tag	Product Name	dark aerobic [SRR097642]	Cassette ID	COG function	KEGG pathway	EC Number	KEGG module
<input type="checkbox"/>	<a href="#">641610012</a>	SYNPCC7002_A0001	chromosomal replication initiator protein DnaA	26.131	<a href="#">1350</a>	<a href="#">L</a> - Replication, recombination and repair	<a href="#">Cell cycle - Caulobacter Two-component system</a>		
<input type="checkbox"/>	<a href="#">641610013</a>	SYNPCC7002_A0002	periplasmic solute-binding protein of ABC transporter	122.720	<a href="#">930</a>	<a href="#">P</a> - Inorganic ion transport and metabolism	<a href="#">ABC transporters</a>		<a href="#">Iron complex transport system</a>
<input type="checkbox"/>	<a href="#">641610014</a>	SYNPCC7002_A0003	hypothetical protein	2898.108	<a href="#">237</a>				
<input type="checkbox"/>	<a href="#">641610015</a>	SYNPCC7002_A0004	hypothetical protein	7474.753	<a href="#">279</a>				
<input type="checkbox"/>	<a href="#">641610016</a>	SYNPCC7002_A0005	hypothetical protein	781.207	<a href="#">255</a>				
<input type="checkbox"/>	<a href="#">641610017</a>	SYNPCC7002_A0006	N-acetylmannosamine-6-phosphate 2-epimerase	89.830	<a href="#">693</a>	<a href="#">G</a> - Carbohydrate transport and metabolism	<a href="#">Amino sugar and nucleotide sugar metabolism</a>	EC:5.1.3.9	

**Figure 5.** Data for Sample



Alternatively, the user can click on “Pathways” (Figure 4) to view all pathways in which genes from this sample are found to participate (Figure 6).

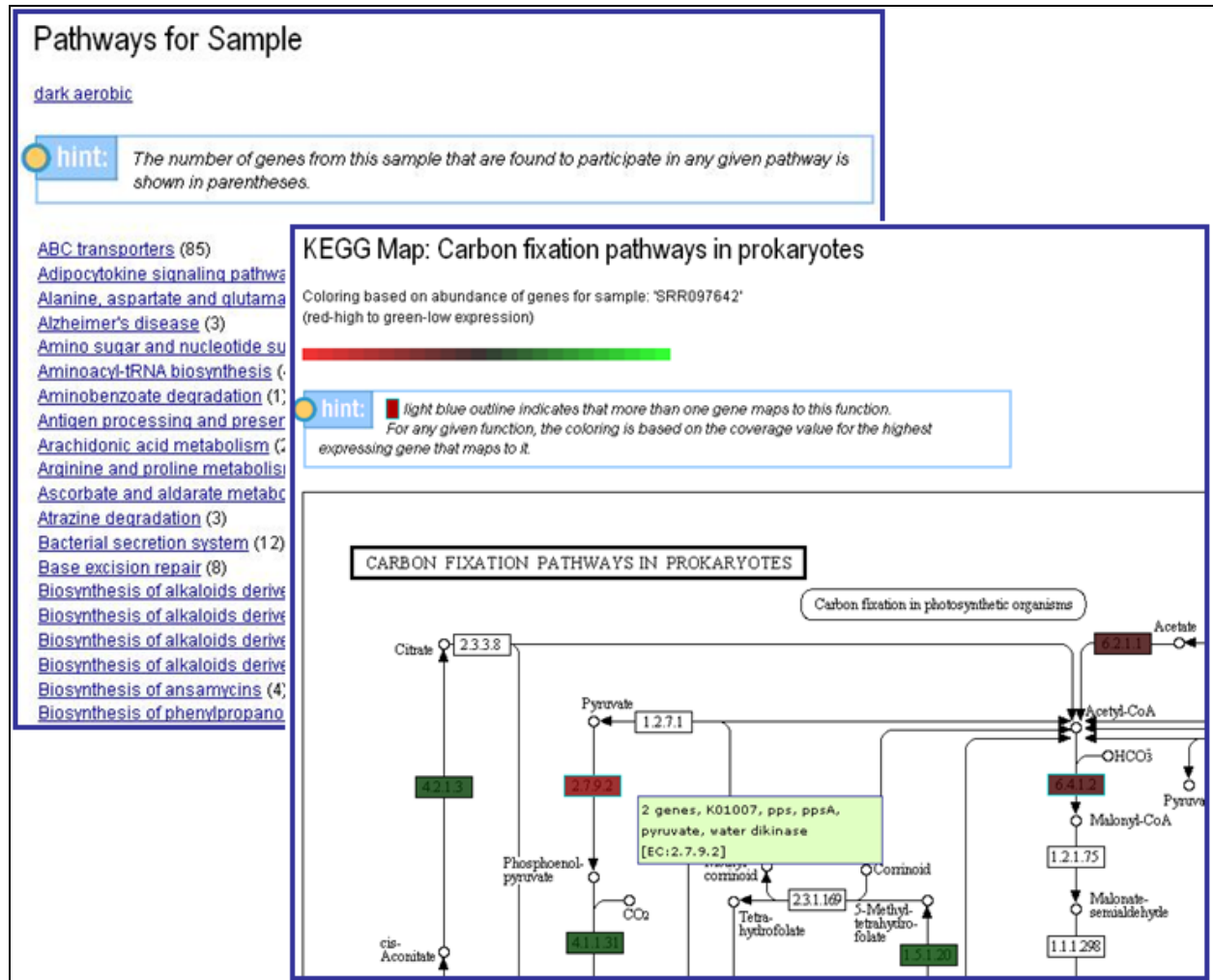


Figure 6. Pathways for a Sample

# KEGG Map: Carbon fixation pathways in prokaryotes

Coloring based on abundance of genes for sample: 'SRR097642'  
(red-high to green-low expression)

**hint:** light blue outline indicates that more than one gene maps to this function.  
For any given function, the coloring is based on the coverage value for the highest expressing gene that maps to it.

**CARBON FIXATION PATHWAYS IN PROKARYOTES**

Carbon fixation in photosynthetic organisms

Key components and pathways shown:

- Citrate** (2.3.3.8) → **Pyruvate** (1.2.7.1) → **Acetyl-CoA** (2.3.1.9)
- Pyruvate** (1.2.7.1) → **Acetyl-CoA** (2.3.1.9)
- Acetyl-CoA** (2.3.1.9) → **Acetyl phosphate** (2.7.2.1) → **Acetate** (2.3.1.8)
- Acetyl-CoA** (2.3.1.9) → **OHCO<sub>3</sub>** (4.1.3.24) → **Pyruvate** (1.2.7.1)
- Acetyl-CoA** (2.3.1.9) → **Malonyl-CoA** (1.2.1.75) → **Malonate-semialdehyde** (1.1.1.298) → **Mesaconyl-C4-CoA** (1.5.1.20) → **Meh** → **(3S)-Citramalyl-CoA** (4.1.3.24) → **Pyruvate** (1.2.7.1)
- Pyruvate** (1.2.7.1) → **Phosphoenolpyruvate** (4.1.1.3) → **cis-Aconitate** (4.2.1.3) → **Citrate** (2.3.3.8)
- Pyruvate** (1.2.7.1) → **2 genes, K01007, pps, ppsA, pyruvate, water dikinase [EC:2.7.9.2]** → **Corrinoid** (2.3.1.169) → **5-Methyl-tetrahydro-folate** (1.5.1.20) → **Tetra-hydrofolate** (2.3.1.169) → **Corrinoid** (2.3.1.169)
- Pyruvate** (1.2.7.1) → **Phosphoenolpyruvate** (4.1.1.3) → **CO<sub>2</sub>** (4.1.1.3) → **Phosphoenolpyruvate** (4.1.1.3)

**Figure 7.** Kegg Map colored by gene expression in single sample

## Pairwise Sample Analysis

The user can select 2 samples of interest to see which genes are up-regulated and which are down-regulated between the two experimental conditions (Figure 8).

### RNASeq Study ?

[Synechococcus sp. PCC 7002](#)

Select Samples

View in GBrowse

Single Sample Analysis

Pairwise Sample Analysis

Multiple Sample Analysis

### Find Up/Down Regulated Genes

You may select 2 samples to identify genes that are up or down regulated.  
Choose [Compare by Function](#) to see whether genes participating in a given function have been up- or down- regulated between the 2 samples as a group.

**Reference:**

☐ high light  
☒ standard1

**Metric:**

☒  $\log R = \log_2(\text{query} / \text{reference})$   
☐  $\text{RelDiff} = 2(\text{query} - \text{reference}) / (\text{query} + \text{reference})$

**Threshold:**  (default=1)

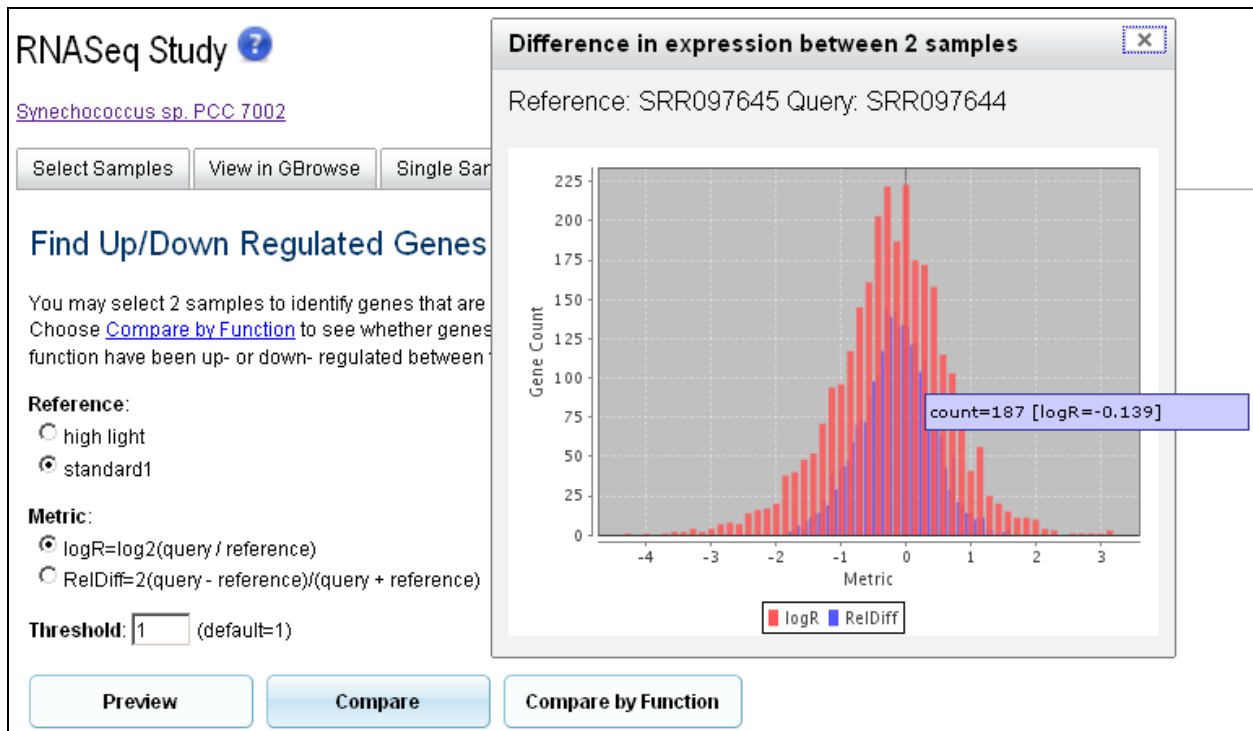
Preview

Compare

Compare by Function

**Figure 8.** Pairwise Sample Analysis

The user can find all the genes whose expression levels differ by a specified threshold. The difference in expression is computed using either the  $\log R = \log_2(\text{query}/\text{reference})$  or the  $\text{RelDiff} = 2(\text{query} - \text{reference}) / (\text{query} + \text{reference})$  metric. The user can preview the comparison to see the value spread in a histogram (Figure 9). This helps in choosing which metric and what threshold value to use.



**Figure 9.** Pairwise Sample Analysis - Preview

When the user clicks “Compare”, the up- and down-regulated genes are presented in separate tabs (Figure 10). The user can then select which genes to add to the Gene Cart.

### Up/Down Regulation

[Transcription profiling of Synechococcus sp. Strain PCC 7002](#)

Reference sample: [standard1](#)  
Query sample: [high light](#)

**hint:** Click on a tab to select up- or down-regulated genes to add to gene cart  
Difference in expression levels is computed using the [logR](#) metric  
Expression levels differ by a [threshold](#): 1  
[Normalization](#): coverage

[Synechococcus sp. PCC 7002](#)

Up-regulated Genes Down-regulated Genes

Add Selections To Gene Cart Select All Clear All

Filter column: [logR](#) Filter term:

Export Page 1 of 2 << first < prev 1 2 next > last >> 100

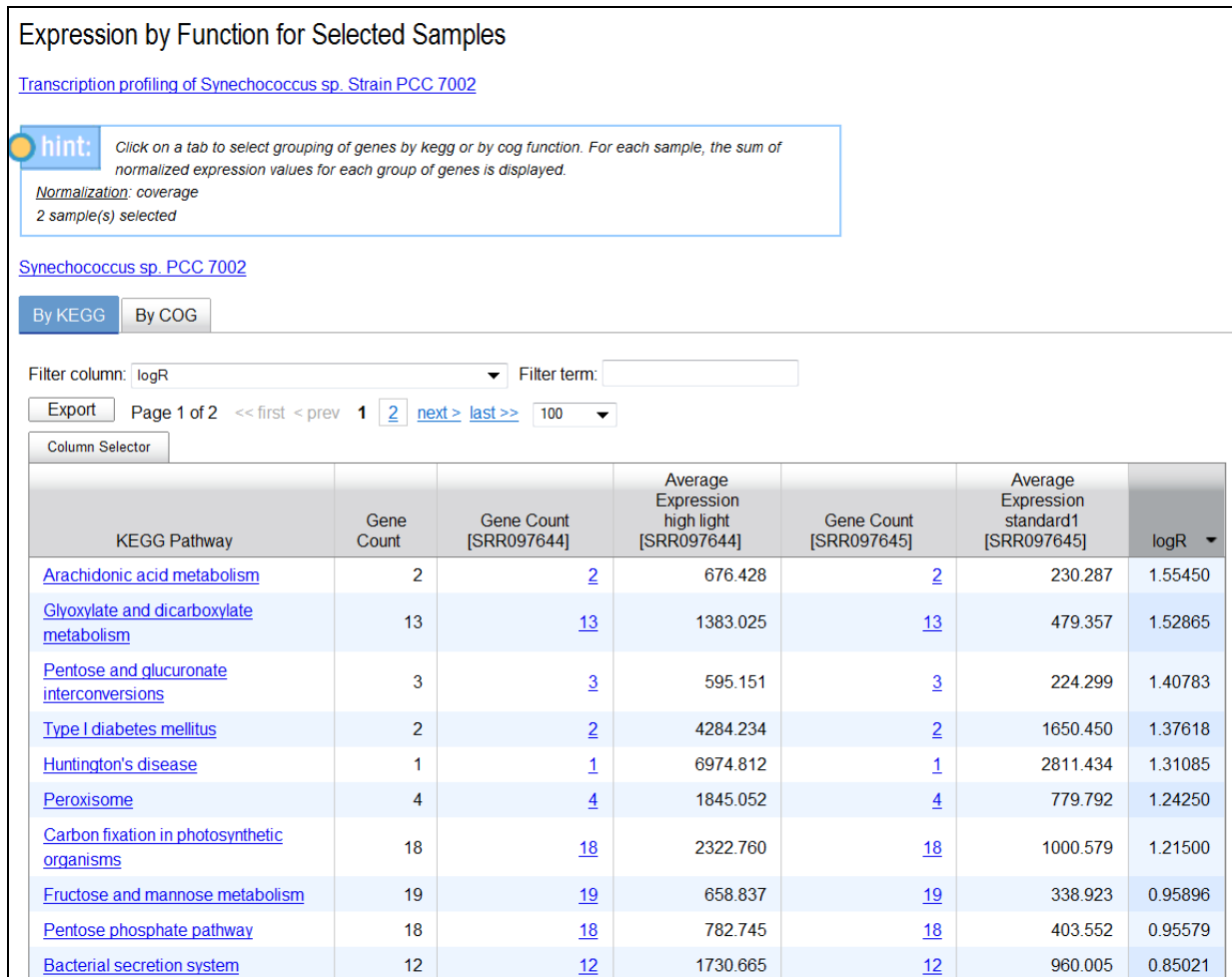
Column Selector Select Page Deselect Page

Select	Gene ID	Locus Tag	Product Name	high light [SRR097644]	standard1 [SRR097645]	logR
<input type="checkbox"/>	<a href="#">641610196</a>	SYNPCC7002_A0186	CAB/ELIP/HLIP superfamily	226.631	2082.948	3.20021
<input type="checkbox"/>	<a href="#">641611983</a>	SYNPCC7002_A1973	NADH dehydrogenase subunit D2	35.033	318.683	3.18534
<input type="checkbox"/>	<a href="#">641611479</a>	SYNPCC7002_A1469	Secreted and surface protein containing fasciclin-like repeats	234.437	2120.555	3.17717
<input type="checkbox"/>	<a href="#">641610169</a>	SYNPCC7002_A0159	omega-3 acyl-lipid desaturase	46.131	361.296	2.96937
<input type="checkbox"/>	<a href="#">641612503</a>	SYNPCC7002_A2493	hypothetical protein	9.074	64.048	2.81934
<input type="checkbox"/>	<a href="#">641611247</a>	SYNPCC7002_A1237	hypothetical protein	20.043	135.812	2.76044

**Figure 10.** Pairwise Sample Analysis - Compare

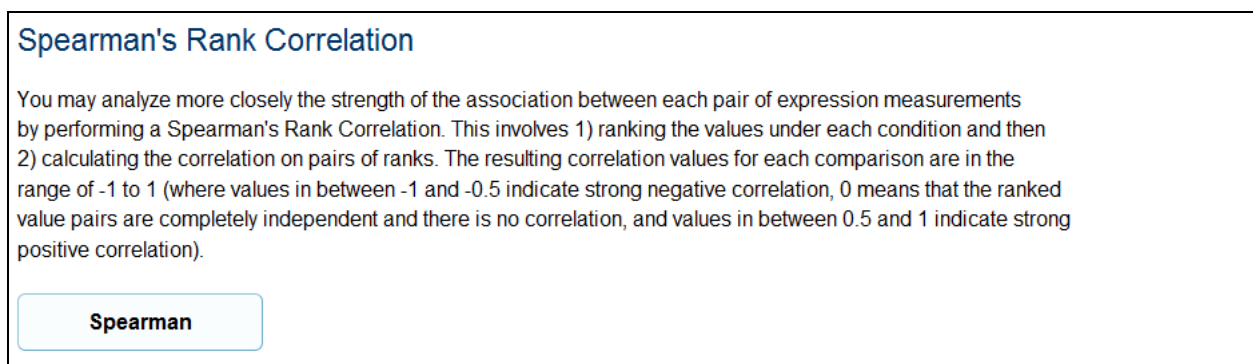


The user can click on “Compare by Function” to see whether genes participating in a given function are up- or down- regulated between the two samples as a group (Figure 11).



**Figure 11.** Pairwise Sample Analysis – Compare by Function

The user can analyze the strength of the association between each pair of expression measurements by performing a Spearman's Rank Correlation (Figure 12).



**Figure 12.** Pairwise Sample Analysis – Spearman's Rank Correlation

## Spearman's Rank Correlation

[Transcription profiling of Synechococcus sp. Strain PCC 7002](#)

Samples compared:

[high light](#)

[standard2](#)

Spearman's Coefficient

Graph

Add Selections To Gene Cart

Select All

Clear All

Filter column: Spearman's Rho

Filter term:

Export

Page 1 of 113 << first < prev

1

2

3

4

5

6

7

8

9

10

next > last >>

25

Column Selector

Select Page

Deselect Page

Select	Gene ID	Locus Tag	Product Name	Spearman's Rho
<input type="checkbox"/>	<a href="#">641610121</a>	SYNPCC7002_A0111	hypothetical protein	1.000
<input type="checkbox"/>	<a href="#">641610243</a>	SYNPCC7002_A0233	Photosystem II subunit PsbJ	1.000
<input type="checkbox"/>	<a href="#">641610483</a>	SYNPCC7002_A0473	hypothetical protein	1.000
<input type="checkbox"/>	<a href="#">641610806</a>	SYNPCC7002_A0796	hypothetical protein	1.000
<input type="checkbox"/>	<a href="#">641611643</a>	SYNPCC7002_A1633	hypothetical protein	1.000
<input type="checkbox"/>	<a href="#">641611819</a>	SYNPCC7002_A1809	hypothetical protein	1.000
<input type="checkbox"/>	<a href="#">641611952</a>	SYNPCC7002_A1942	hypothetical protein	1.000

## Spearman's Rank Correlation

[Transcription profiling of Synechococcus sp. Strain PCC 7002](#)

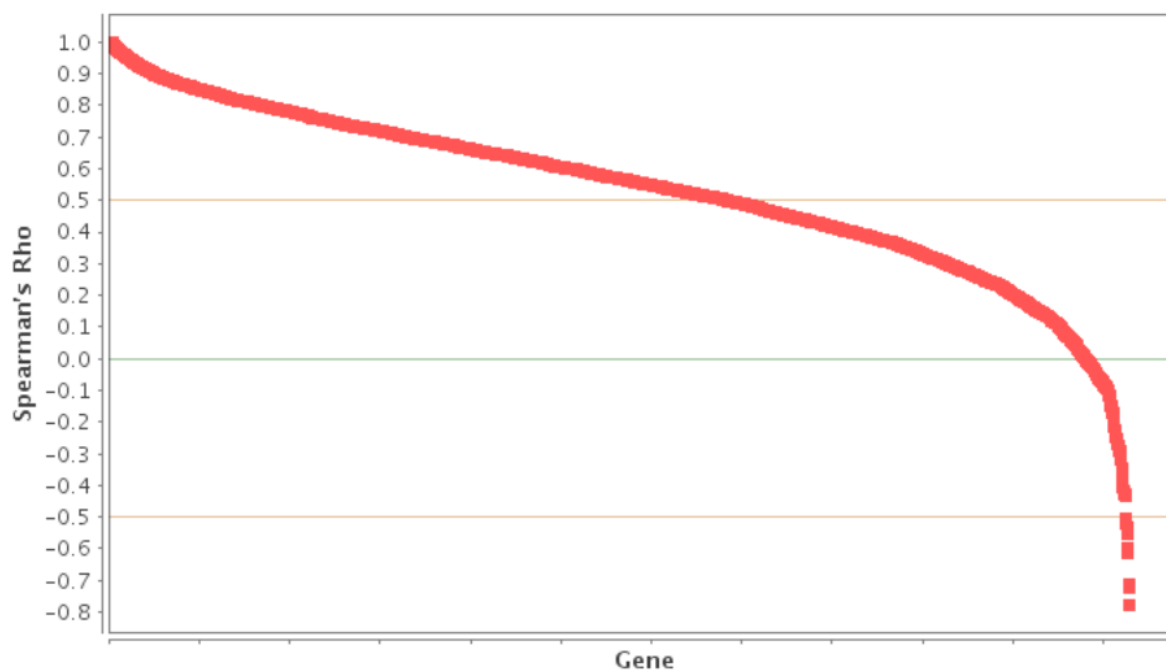
Samples compared:

[high light](#)

[standard2](#)

Spearman's Coefficient

Graph



## Multiple Sample Analysis

Multiple samples can be analyzed to show gene coverage and to display information and links to cassettes, associated cog functions, kegg pathways and modules, and EC (Figure 14). This information is displayed when the user selects samples and clicks on “Gene Expression Summary” (Figure 13).

### RNASeq Study ?

[Synechococcus sp. PCC 7002](#)

[Select Samples](#) [View in GBrowse](#) [Single Sample Analysis](#) [Pairwise Sample Analysis](#) [Multiple Sample Analysis](#)

#### Describe Samples

You may select samples to compare and describe. If you choose to first cluster the samples to have the results colored by the cluster grouping, go to [Map Clusters to Pathways](#) below. Choose [Expression by Function](#) to see, for each selected sample, the average expression of genes participating in a given function as a group.

☐ Use only genes from gene cart

[Gene Expression Summary](#) [Expression by Function](#)

#### Cluster Samples

You may select samples and cluster them based on the abundance of the expressed genes. Proximity of grouping indicates the relative degree of similarity of samples to each other.

**Figure 13.** Multiple Sample Analysis

The samples can be clustered before displaying gene expression summary. This way the results will be colored by the cluster groupings (see below under “Sample Clustering”).

RNASeq Expression Data for Selected Samples

5 sample(s) selected  
Normalization: coverage

Transcription profiling of *Synechococcus* sp. Strain PCC 7002

Add Selections To Gene Cart

Select All

Clear All

Filter column: Gene ID Filter term:

Export

Page 1 of 29

<< first < prev

1

2

3

4

5

6

7

8

9

10

next > last >>

100

Column Selector

Select Page

Deselect Page

Select	Gene ID	Locus Tag	Product Name	dark aerobic [SRR097642]	dark anaerobic [SRR097643]	high light [SRR097644]	standard1 [SRR097645]	low O2 [SRR097656]	Cassette ID	COG function	KEGG pathway	EC Number	KEGG module
<input type="checkbox"/>	<a href="#">641610012</a>	SYNPCC7002_A0001	chromosomal replication initiator protein DnaA	26.131	20.076	48.534	45.101	34.498	<a href="#">1350</a>	L - Replication, recombination and repair	<a href="#">Cell cycle - Caulobacter Two-component system</a>		
<input type="checkbox"/>	<a href="#">641610013</a>	SYNPCC7002_A0002	periplasmic solute-binding protein of ABC transporter	122.720	37.468	36.363	79.064	52.714	<a href="#">930</a>	P - Inorganic ion transport and metabolism	<a href="#">ABC transporters</a>		<a href="#">Iron complex transport system</a>
<input type="checkbox"/>	<a href="#">641610014</a>	SYNPCC7002_A0003	hypothetical protein	2898.108	6334.413	151.607	474.504	341.308	<a href="#">237</a>				
<input type="checkbox"/>	<a href="#">641610015</a>	SYNPCC7002_A0004	hypothetical protein	7474.753	7951.578	128.784	446.004	614.999	<a href="#">279</a>				
<input type="checkbox"/>	<a href="#">641610016</a>	SYNPCC7002_A0005	hypothetical protein	781.207	516.228	82.885	101.771	105.738	<a href="#">255</a>				
<input type="checkbox"/>	<a href="#">641610017</a>	SYNPCC7002_A0006	N-acetylmannosamine-6-phosphate 2-epimerase	89.830	30.728	15.249	35.048	35.371	<a href="#">683</a>	G - Carbohydrate transport and metabolism	<a href="#">Amino sugar and nucleotide sugar metabolism</a>	EC:5.1.3.9	

**Figure 14.** Data for Samples

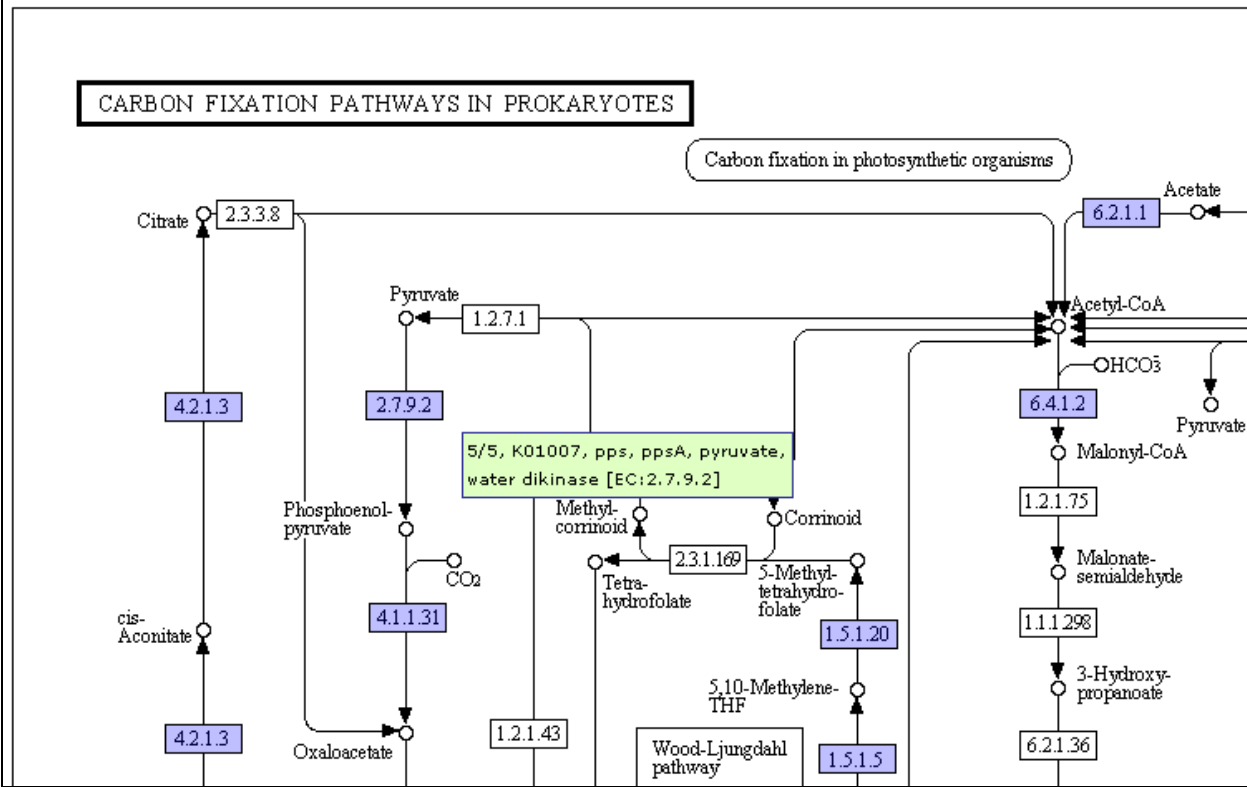
The kegg pathway link from the “expression data” page, displays the pathway for multiple samples that the user selected. The pathway will be colored by percentage of samples in which a given gene is expressed (Figure 15).

## KEGG Map: Carbon fixation pathways in prokaryotes

5 samples selected ('SRR097642','SRR097643','SRR097644','SRR097645','SRR097656')

■ Genes found in all selected samples

■ Genes found in some of the selected samples [for up to 25% ■ >25% ■ >50% ■ >75% ■]



**Figure 15.** Pathway for Samples (unclustered)

Each region of interest on the kegg map links to a list of genes for the selected samples that fall in that region (Figure 16).

## Genes with KO: [K01682](#) found among selected samples

Add Selected to Gene Cart

Select All

Clear All

Filter column: 

Gene Object ID

Filter term:

Export

Page 1 of 1

<< first < prev 1 next > last >>

All

Column Selector

Select Page

Deselect Page

Select	Gene Object ID	Locus Tag	Gene Product Name	Sample ID	Sample Description	KO
<input type="checkbox"/>	<a href="#">641611693</a>	SYNPCC7002_A1683	aconitate hydratase 2	SRR097642	<a href="#">dark aerobic</a>	<a href="#">K01682</a>
<input type="checkbox"/>	<a href="#">641611693</a>	SYNPCC7002_A1683	aconitate hydratase 2	SRR097643	<a href="#">dark anaerobic</a>	<a href="#">K01682</a>
<input type="checkbox"/>	<a href="#">641611693</a>	SYNPCC7002_A1683	aconitate hydratase 2	SRR097644	<a href="#">high light</a>	<a href="#">K01682</a>
<input type="checkbox"/>	<a href="#">641611693</a>	SYNPCC7002_A1683	aconitate hydratase 2	SRR097645	<a href="#">standard1</a>	<a href="#">K01682</a>
<input type="checkbox"/>	<a href="#">641611693</a>	SYNPCC7002_A1683	aconitate hydratase 2	SRR097656	<a href="#">low O2</a>	<a href="#">K01682</a>

Export

Page 1 of 1

<< first < prev 1 next > last >>

All

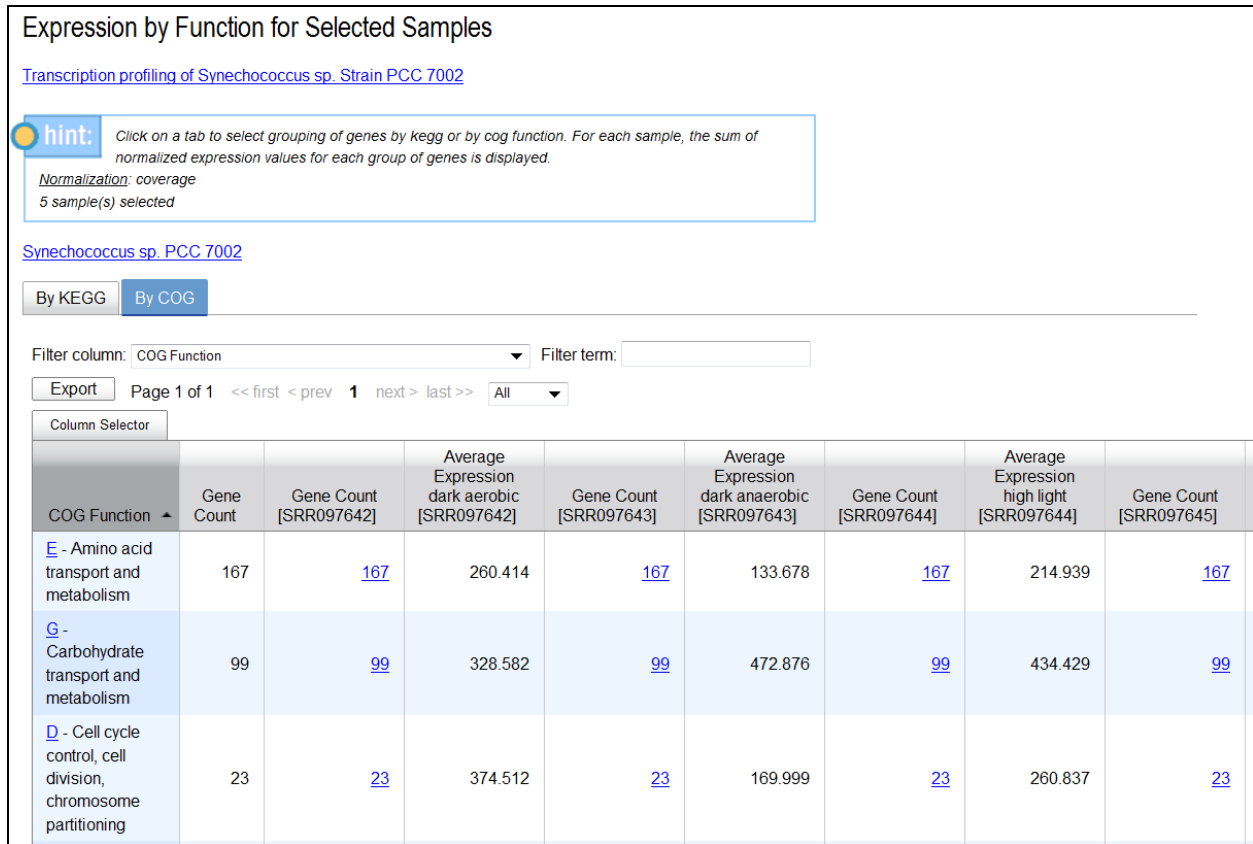
Add Selected to Gene Cart

Select All

Clear All

**Figure 16.** List of genes for selected samples that fall on a given region of interest

On the “Multiple Sample Analysis” page, the user can also click on “Expression by Function” (Figure 13) to see, for each selected sample, the average expression of genes participating in a given function as a group (Figure 17).



**Figure 17.** Multiple Sample Analysis – Expression by Function

## Sample Clustering

For each study, the experiments (samples) can be selected for **clustering** based on coverage values for the genes expressed in each sample (Figure 18). The user can specify the clustering method and the distance measure to use, as well as the minimum number of samples in which a gene should appear in order to be included in the calculation.

### Cluster Samples

You may select samples and cluster them based on the abundance of the expressed genes. Proximity of grouping indicates the relative degree of similarity of samples to each other.

**Clustering Method:**

☒ Pairwise complete-linkage (default)  
☐ Pairwise single-linkage  
☐ Pairwise average-linkage

**Distance Measure:**

☒ Pearson correlation (default)  
☐ Spearman's rank correlation  
☐ Euclidean distance  
☐ City-block distance (Manhattan)

Minimum number of samples in which a gene should appear in order to be included:  (default=2)

Cut-off threshold:  (default=0.8)

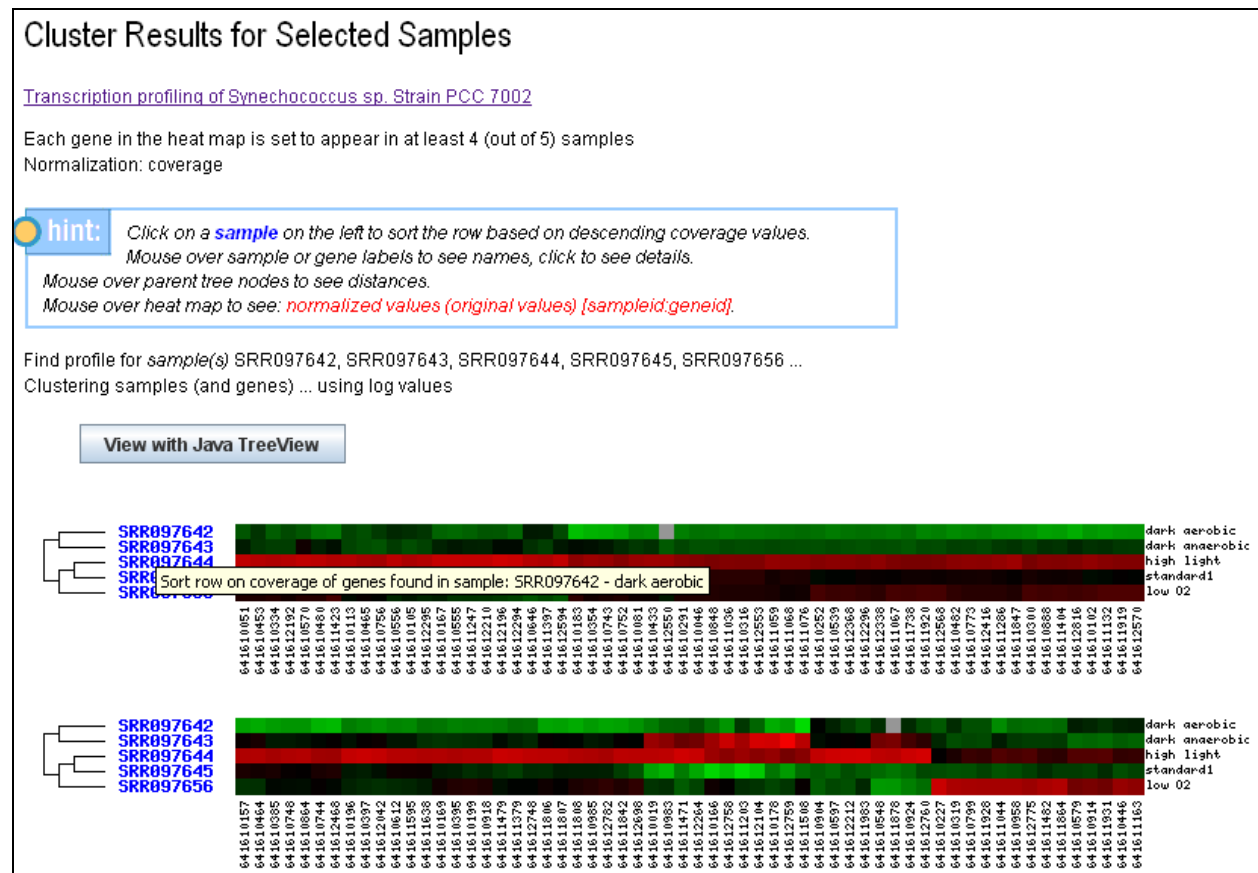
☐ Use only genes from gene cart

ClusterMap Clusters to Pathways

**Figure 18.** Clustering of rnaseq samples

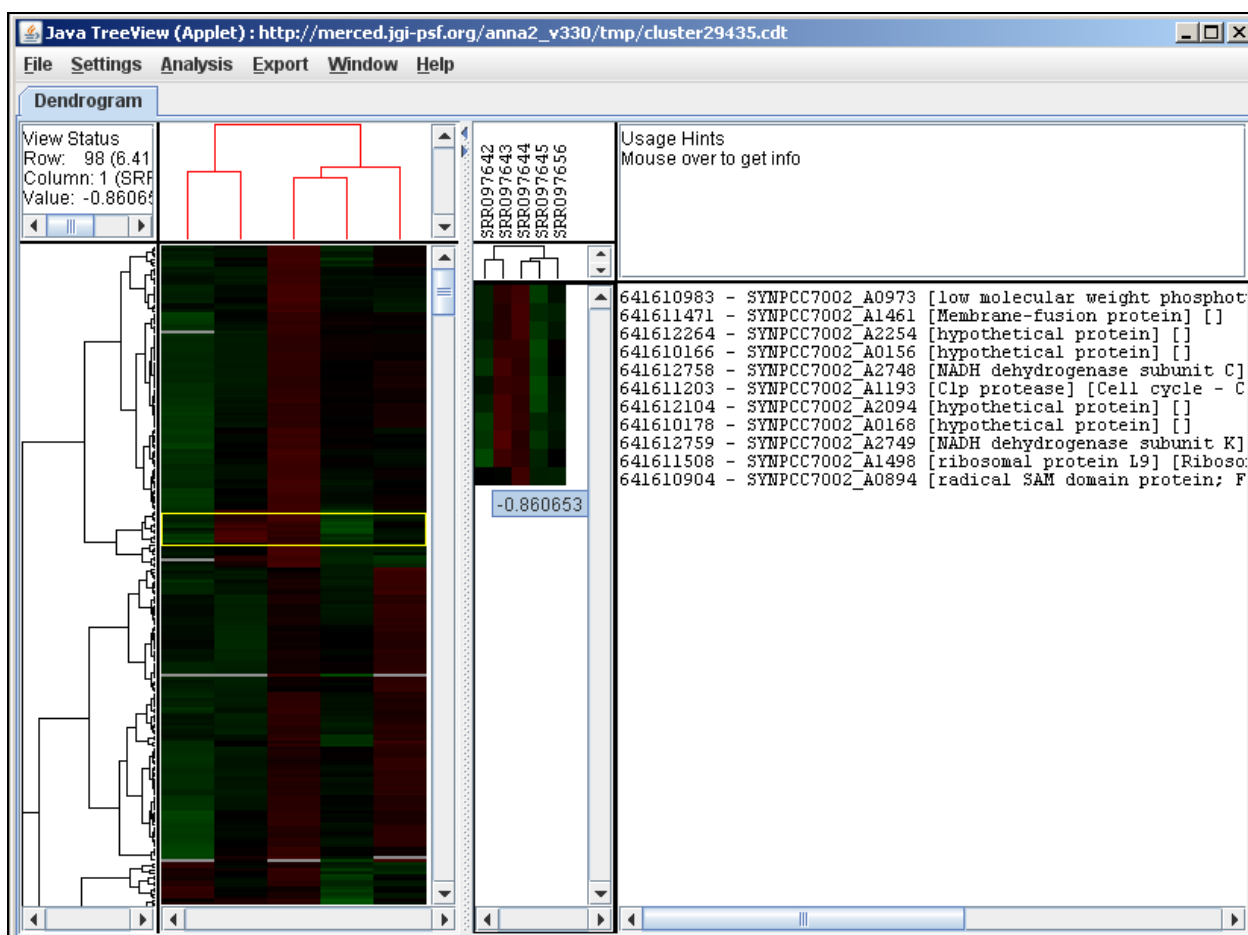


Both samples and genes are clustered. What will be displayed in IMG are a hierarchical cluster tree of samples and a normalized heat map of coverage values for each gene for each sample (Figure 19). On the heat map, the samples link to sample pages and the genes link to gene pages. Clicking on a sample in the cluster tree **sorts** the heat map on descending coverage values for that sample.



**Figure 19.** Clustering results

In addition, the results can be viewed in **Java TreeView** applet (Figure 20). The applet is loaded with the cluster files by clicking on the “View with Java TreeView” button. Within the Java TreeView, the genes link to IMG gene pages and the samples link to IMG sample pages.



**Figure 20.** Cluster results displayed in Java TreeView application

## Sample Clustering

Multiple samples can be clustered to create groupings of genes. Expression data for the samples can then be sorted by these cluster groups (Figure 21). This is done when the user clicks on “Map Clusters to Pathways” under the “Multiple Sample Analysis” tab (Figure 18).

RNASeq Expression Data for Selected Samples

5 sample(s) selected  
Normalization: coverage

Transcription profiling of *Synechococcus* sp. Strain PCC 7002

Note: selecting too few conditions may result in skewed clusters.  
Clustering samples (and genes) ... using log values  
Cut-off threshold set to: 0.8

Add Selections To Gene Cart Select All Clear All

Filter column: Gene ID Filter term:

Export Page 1 of 29 << first < prev 1 2 3 4 5 6 7 8 9 10 next > last >> 100

Column Selector Select Page Deselect Page

Select	Gene ID	Locus Tag	Product Name	Cluster ID	dark aerobic [SRR097642]	dark anaerobic [SRR097643]	high light [SRR097644]	standard1 [SRR097645]	low O2 [SRR097656]	Cassette ID	COG function	KEGG pathway	EC Number	KEGG module
<input type="checkbox"/>	641610012	SYNPPC7002_A0001	chromosomal replication initiator protein DnaA	5	26.131	20.076	48.534	45.101	34.498	1350	L - Replication, recombination and repair	Cell cycle - Caulobacter Two-component system		
<input type="checkbox"/>	641610013	SYNPPC7002_A0002	periplasmic solute-binding protein of ABC transporter	79	122.720	37.468	36.363	79.064	52.714	930	P - Inorganic ion transport and metabolism	ABC transporters		Iron complex transport system
<input type="checkbox"/>	641610014	SYNPPC7002_A0003	hypothetical protein	74	2898.108	6334.413	151.607	474.504	341.308	237				
<input type="checkbox"/>	641610015	SYNPPC7002_A0004	hypothetical protein	73	7474.753	7951.578	128.784	446.004	614.999	279				
<input type="checkbox"/>	641610016	SYNPPC7002_A0005	hypothetical protein	72	781.207	516.228	82.885	101.771	105.738	255				
<input type="checkbox"/>	641610017	SYNPPC7002_A0006	N-acetylmannosamine-6-phosphate 2-epimerase	80	89.830	30.728	15.249	35.048	35.371	693	G - Carbohydrate transport and metabolism	Amino sugar and nucleotide sugar metabolism	EC:5.1.3.9	

**Figure 21.** Map clusters to pathways

The kegg pathway link from the sample “expression data” page, displays the pathway colored by the cluster groupings (Figure 22).

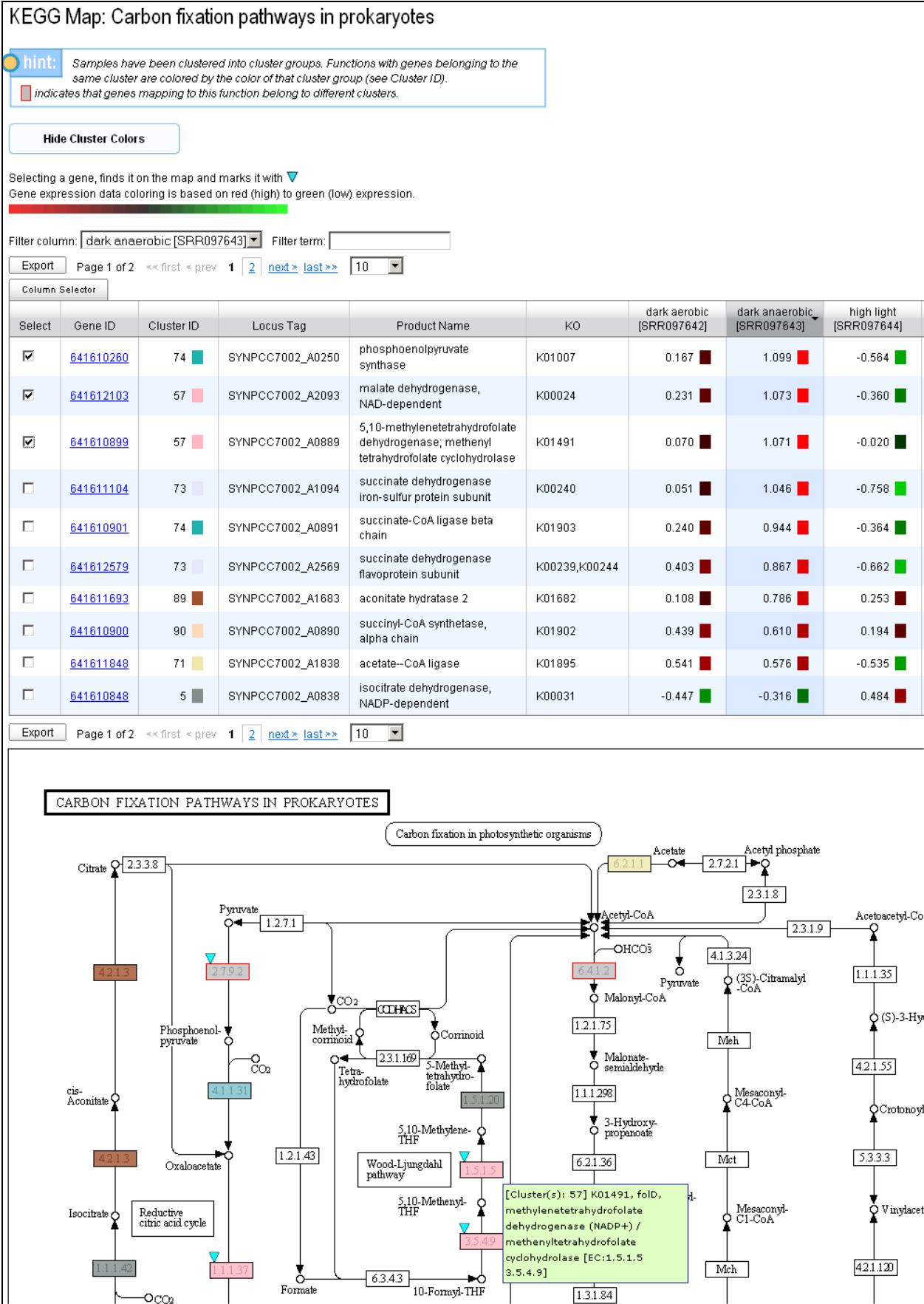


Figure 22. Multiple Sample Analysis – Map Clusters to Pathway

## View in GBrowse

The samples for an rnaseq study can also be analyzed in GBrowse to view differences in read coverage for individual genes under various conditions (Figure 23).

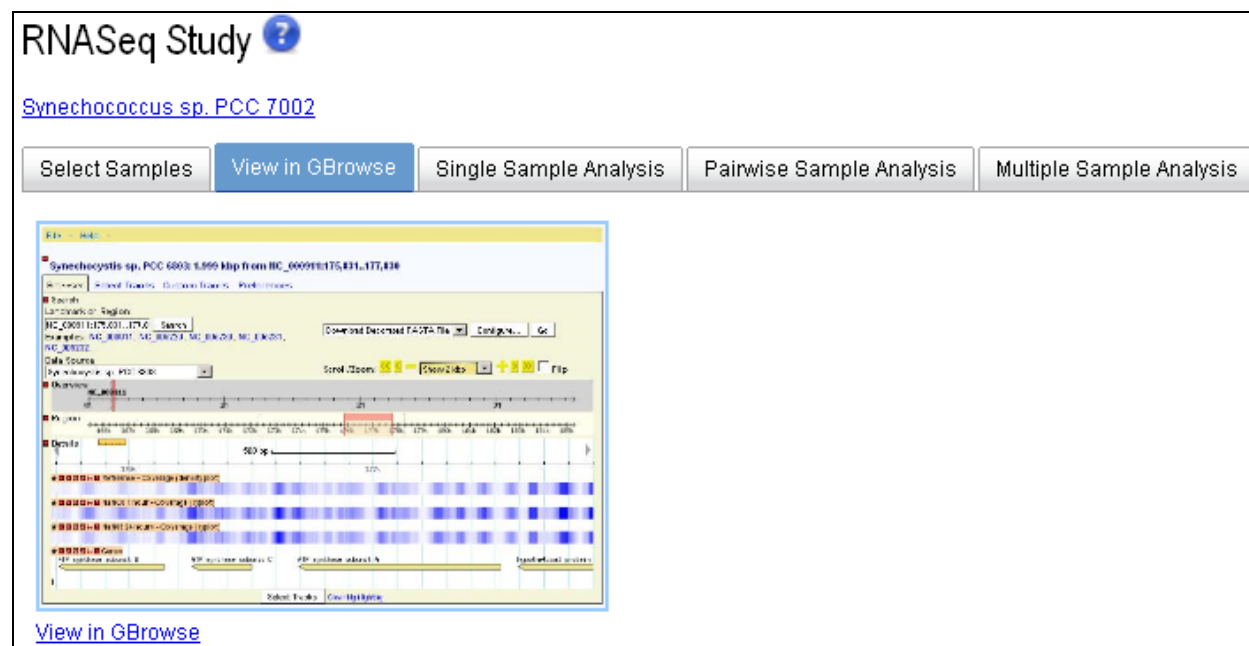


Figure 23. View in GBrowse

Clicking on “View in GBrowse” link or on the image of GBrowse will open the data for this study in GBrowse (Figure 24).

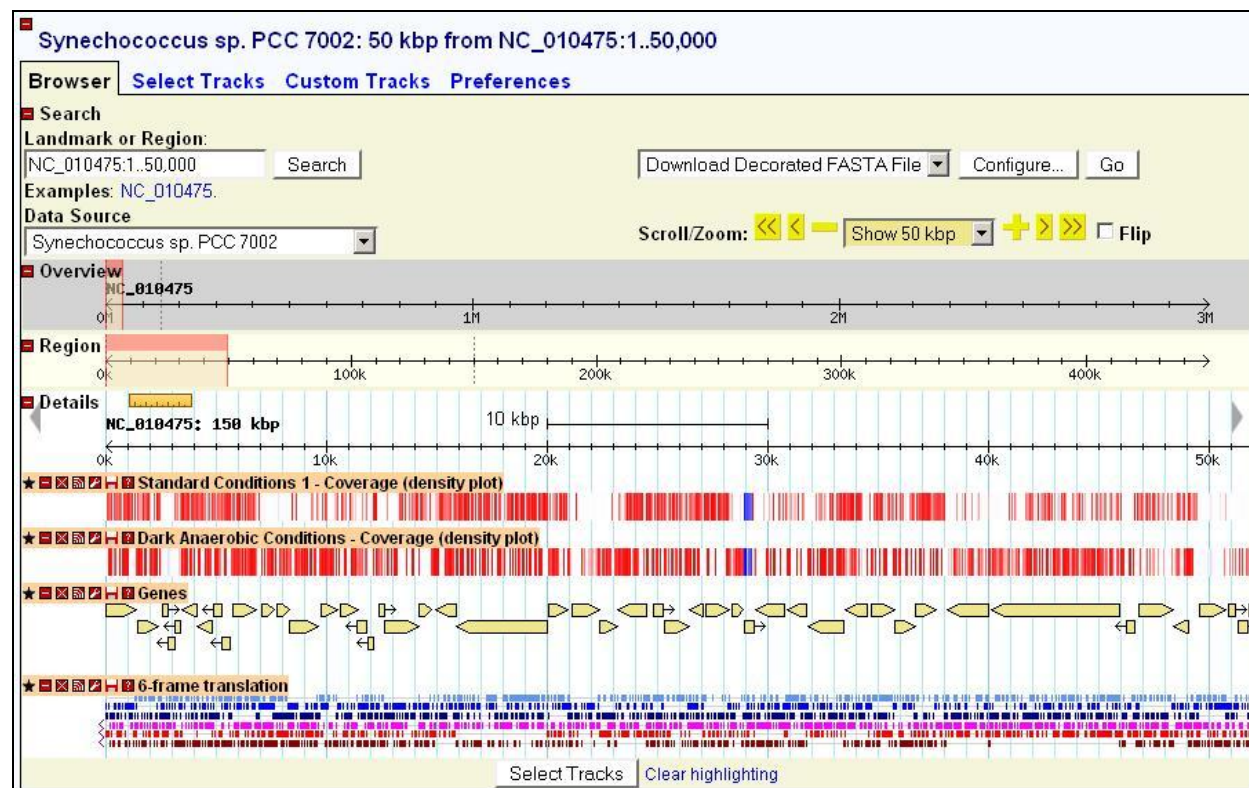


Figure 24. Study loaded into GBrowse

After the data is loaded, the user needs to click on “Select Tracks” to select which data to display (Figure 25).

**Synechococcus sp. PCC 7002: 50 kbp from NC\_010475:1..50,000**

Browser **Select Tracks** Custom Tracks Preferences

<< Back to Browser Show Favorites Only ★ Clear All Favorites ★

Tracks

Experiments ☐ All on ☐ All off

CoverageXYPlot	CoverageDensity	Reads
SRR097642 <input type="checkbox"/> Dark Aerobic Conditions - Coverage (xyplot)	<input type="checkbox"/> Dark Aerobic Conditions - Coverage (density plot)	<input type="checkbox"/> Dark Aerobic - Reads
SRR097643 <input type="checkbox"/> Dark Anaerobic Conditions - Coverage (xyplot)	<input checked="" type="checkbox"/> Dark Anaerobic Conditions - Coverage (density plot)	<input type="checkbox"/> Dark Anaerobic - Reads
SRR097644 <input type="checkbox"/> High Light Exposure - Coverage (xyplot)	<input type="checkbox"/> High Light Exposure - Coverage (density plot)	<input type="checkbox"/> High Light Exposure - Reads
SRR097645 <input type="checkbox"/> Standard Conditions 1 - Coverage (xyplot)	<input checked="" type="checkbox"/> Standard Conditions 1 - Coverage (density plot)	<input type="checkbox"/> Standard Conditions 1 - Reads
SRR097646 <input type="checkbox"/> Standard Conditions 2 - Coverage (xyplot)	<input type="checkbox"/> Standard Conditions 2 - Coverage (density plot)	<input type="checkbox"/> Standard Conditions 2 - Reads
SRR097647 <input type="checkbox"/> Standard Conditions 3 - Coverage (xyplot)	<input type="checkbox"/> Standard Conditions 3 - Coverage (density plot)	<input type="checkbox"/> Standard Conditions 3 - Reads
SRR097648 <input type="checkbox"/> Cell Density 0.4 - Coverage (xyplot)	<input type="checkbox"/> Cell Density 0.4 - Coverage (density plot)	<input type="checkbox"/> Cell Density 0.4 - Reads
SRR097649 <input type="checkbox"/> Cell Density 1.0 - Coverage (xyplot)	<input type="checkbox"/> Cell Density 1.0 - Coverage (density plot)	<input type="checkbox"/> Cell Density 1.0 - Reads
SRR097650 <input type="checkbox"/> Cell Density 3.0 - Coverage (xyplot)	<input type="checkbox"/> Cell Density 3.0 - Coverage (density plot)	<input type="checkbox"/> Cell Density 3.0 - Reads
SRR097651 <input type="checkbox"/> Cell Density 5.0 - Coverage (xyplot)	<input type="checkbox"/> Cell Density 5.0 - Coverage (density plot)	<input type="checkbox"/> Cell Density 5.0 - Reads
SRR097656 <input type="checkbox"/> low O2 - Coverage (xyplot)	<input type="checkbox"/> low O2 - Coverage (density plot)	<input type="checkbox"/> low O2 - Reads

Scaffolds

DNA ☐ All on ☐ All off

☐ DNA/GC Content

Genes ☐ All on ☐ All off

☒ Genes

Translation ☐ All on ☐ All off

☒ 6-frame translation

Back to Browser

**Figure 25.** GBrowse – Select Tracks

The user can also zoom closer on the chromosome and see the actual DNA and nucleotide sequences for the genes.

**Synechococcus sp. PCC 7002: 999 bp from NC\_010475:106,872..107,871**

Browser **Select Tracks** Custom Tracks Preferences

Search

Landmark or Region: NC\_010475:106,872..107,871 Search

Download Decorated FASTA File Configure... Go

Examples: NC\_010475.

Data Source: Synechococcus sp. PCC 7002

Scroll/Zoom: << < > >> Show 1 kbp + > >> Flip

Overview

Region

Details

Standard Conditions 1 - Coverage (density plot)

Dark Anaerobic Conditions - Coverage (density plot)

Genes

glyceraldehyde-3-phosphate dehydrogenase

6-frame translation

Select Tracks Clear highlighting

**Figure 26.** GBrowse – Zoom



## Chromosome Viewer – Color by “Expression”

There are 2 ways to color the chromosome by expression.

From the [sample page](#) - from each experimental sample page, the user can select to go to the Chromosome Viewer page (Figure 27).

RNASeq Expression Data for Selected Sample

[dark aerobic](#)

**Chromosome Viewer**

[Transcription profiling of \*Synechococcus\* sp. Strain PCC 7002](#)

[Add Selections To Gene Cart](#) [Select All](#) [Clear All](#)

Filter column:  Filter term:

[Export](#) Page 1 of 29 << first < prev 1 2 3 4 5 6 7 8 9 10 next > last >> 100

Column Selector [Select Page](#) [Deselect Page](#)

Select	Gene ID	Locus Tag	Product Name	dark aerobic [SRR097642]	Cassette ID	COG function	KEGG pathway	EC Number	KEGG module
<input type="checkbox"/>	<a href="#">641610012</a>	SYNPCC7002_A0001	chromosomal replication initiator protein DnaA	26.131	<a href="#">1350</a>	L - Replication, recombination and repair	<a href="#">Cell cycle - Caulobacter Two-component system</a>		
<input type="checkbox"/>	<a href="#">641610013</a>	SYNPCC7002_A0002	periplasmic solute-binding protein of ABC transporter	122.720	<a href="#">930</a>	P - Inorganic ion transport and metabolism	<a href="#">ABC transporters</a>		<a href="#">Iron complex transport system</a>
<input type="checkbox"/>	<a href="#">641610014</a>	SYNPCC7002_A0003	hypothetical protein	2898.108	<a href="#">237</a>				
<input type="checkbox"/>	<a href="#">641610015</a>	SYNPCC7002_A0004	hypothetical protein	7474.753	<a href="#">279</a>				
<input type="checkbox"/>	<a href="#">641610016</a>	SYNPCC7002_A0005	hypothetical protein	781.207	<a href="#">255</a>				
<input type="checkbox"/>	<a href="#">641610017</a>	SYNPCC7002_A0006	N-acetylmannosamine-6-phosphate 2-epimerase	89.830	<a href="#">693</a>	G - Carbohydrate transport and metabolism	<a href="#">Amino sugar and nucleotide sugar metabolism</a>	EC:5.1.3.9	

**Figure 27.** Chromosome Viewer link from Sample page

From the [genome page](#) - if a genome has samples associated with it, the “Scaffolds and Contigs” page allows the user to display the list of all available samples (by clicking on the “Select RNASeq Samples” button – Figure 28) and to choose one sample to color the chromosome by the expression of genes for that sample.

Chromosome Viewer

Scaffolds and contigs for [Synechococcus sp. PCC 7002](#)

**Select RNASeq Samples**

User Selectable Coordinates

[Add to Scaffold Cart](#) [Select All](#) [Clear All](#)

Filter column:  Filter term:

[Export](#) Page 1 of 1 << first < prev 1 next > last >> All

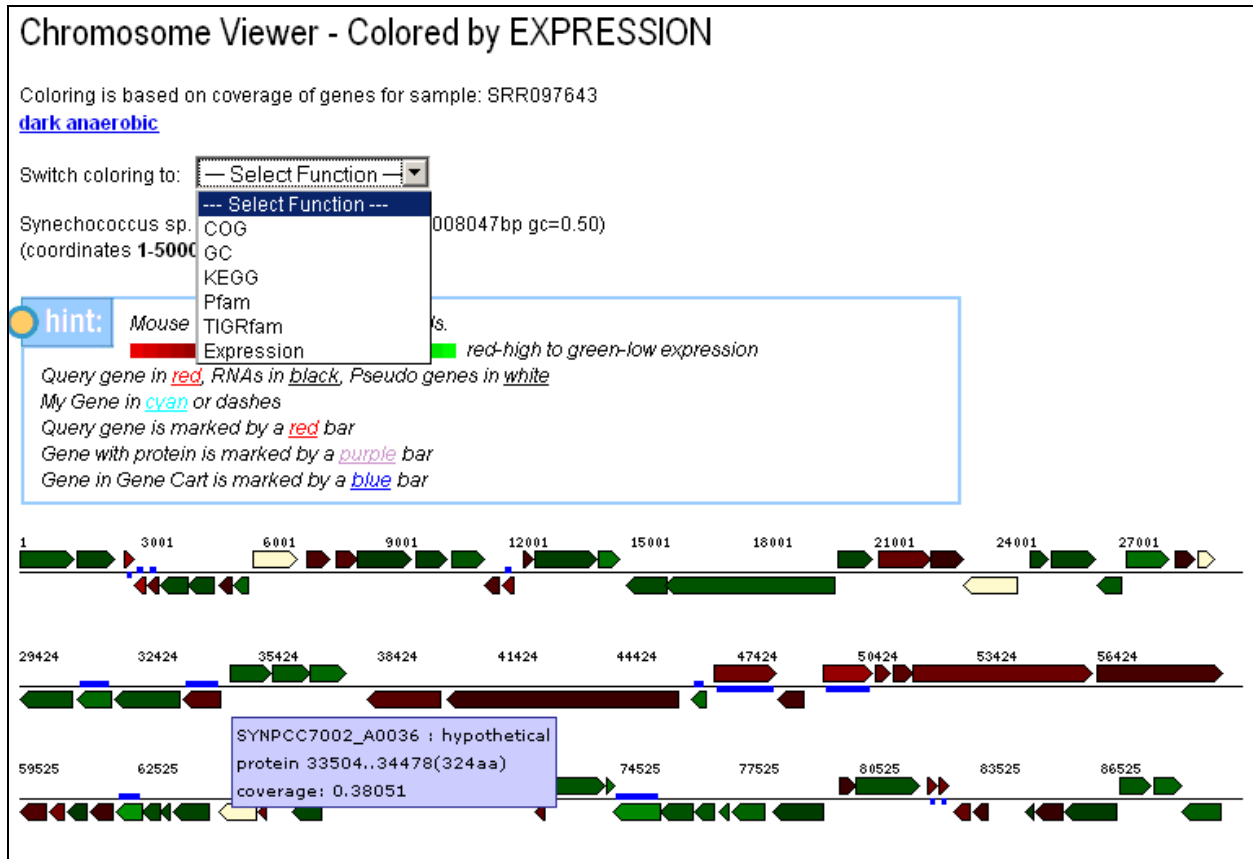
Column Selector [Select Page](#) [Deselect Page](#)

Select	Scaffold	Length (bp)	GC	Type	Topology	No. Genes	Coordinate Range
<input type="checkbox"/>	Synechococcus sp. PCC 7002 plasmid pAQ1: NC_010476	4809	0.49	plasmid	circular	3	<a href="#">1_4809</a>
<input type="checkbox"/>	Synechococcus sp. PCC 7002 plasmid pAQ3: NC_010477	16103	0.46	plasmid	circular	17	<a href="#">1_16103</a>
<input type="checkbox"/>	Synechococcus sp. PCC 7002 plasmid pAQ4: NC_010478	31972	0.44	plasmid	circular	30	<a href="#">1_31972</a>
<input type="checkbox"/>	Synechococcus sp. PCC 7002 plasmid pAQ5: NC_010479	38515	0.43	plasmid	circular	39	<a href="#">1_38515</a>

**Figure 28.** User can choose samples to use for coloring the chromosome viewer



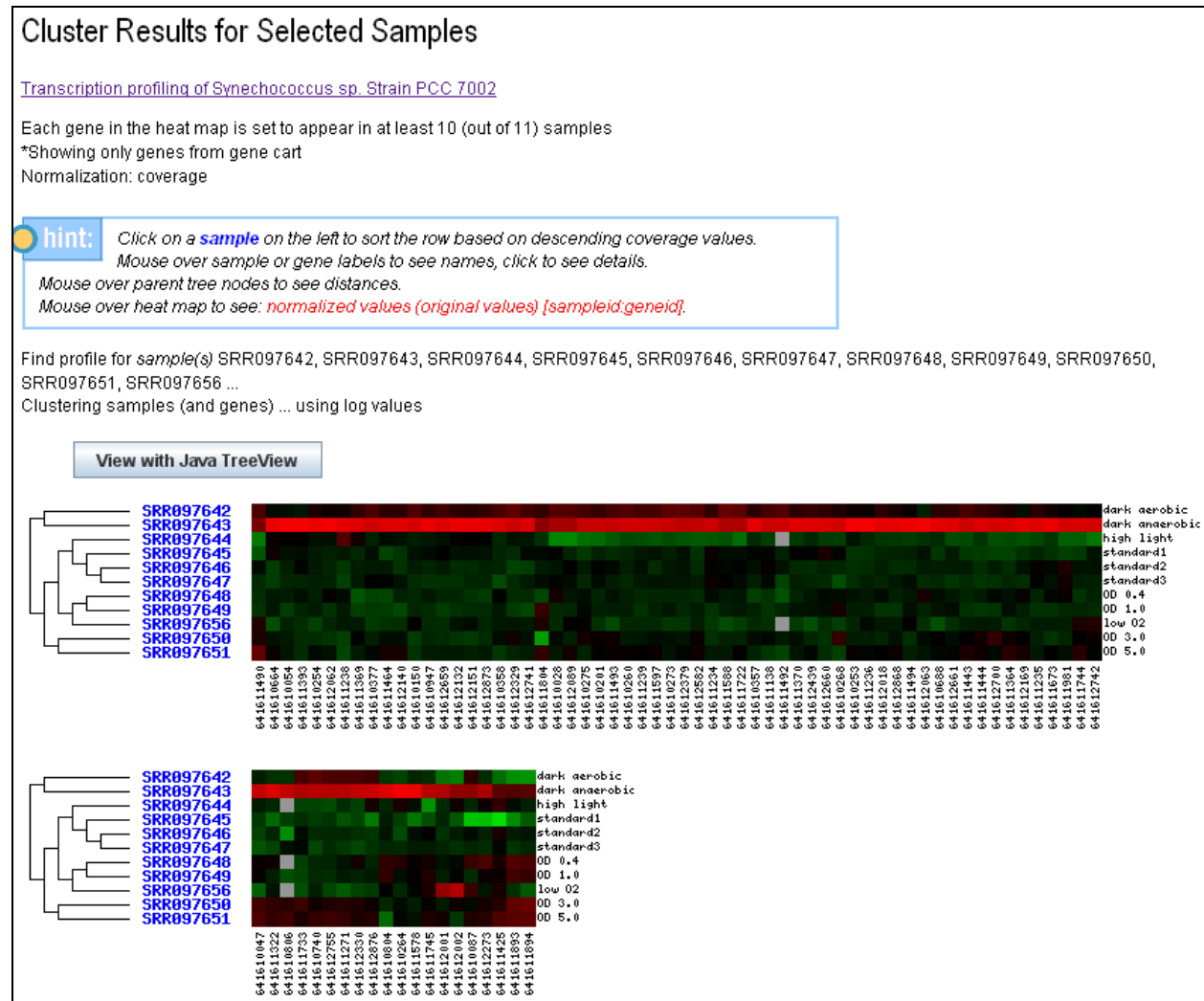
The Chromosome Viewer that opens up displays coloring by “expression”, that is, coverage values for each gene in the given sample (Figure 29).



**Figure 29.** Chromosome Viewer colored by gene expression in a given sample

## Analyzing genes of interest

The user can choose to limit the analysis to only the genes that were added to the Gene Cart. For example, after a pairwise analysis the user can add the top 100 genes, showing the greatest logR value, to the Gene Cart. The user can go back and select all samples. Then, returning to the “Cluster Samples” page under “Multiple Sample Analysis” tab (Figure 13), the user can check the “Use only genes from gene cart” and then click “Cluster” (Figure 18). Alternatively, the user can use “Gene Expression Summary” to see the Gene Cart genes in the context of the KEGG maps, chromosomal neighborhood, and cassettes that they belong to.



**Figure 30.** Clustering results – using only genes from Gene Cart